

How will it end?

Predicting Cancer Clinical Trial Completion

Team Pioneer: Aziz Burak Guelen, Kriti Sehgal, Asia Wyatt
Spring 2022 Erdős Institute Data Science Bootcamp

The clinical trial process is a multi-year, often decade-long process that involves rigorous research and testing from the petri dish to double-blind human tests. The overall cost to bring a single intervention to market can cost in the range of hundreds of millions of dollars to well over a billion dollars. Once a drug reaches the clinical trial stage, there are many factors that are taken into consideration for how to structure said trial—the age and number of participants, levels of masking, the location for the trial, etc.

While the goal is always to reach completion, there are a myriad of reasons why a clinical trial may be terminated or suspended. These can range from adverse side effects and poor safety to funding setbacks and lack of efficacy. Given that it is advantageous for investors and regulators to create a clear pipeline to drug development with the highest likelihood for success, we see benefit in analyzing the ways in which these studies are structured.

The goal of our project is to provide insight into which of these factors are most important in having a clinical trial make it to completion. Specifically, we aim to determine which of these factors when structuring a clinical trial are the most important for cancer intervention trial completion and to predict whether or not a test set of trials will be completed. Using data provided by <https://clinicaltrials.gov/> we assessed interventional cancer studies that aimed to bring a new treatment to the market.

Given our initial research and intuition on viability of clinical trials we selected to analyze the following published properties of our subset of clinical trials:

- Conditions (cancer types the therapy is tested for)
- Funder Type (NIH/Federal funding, Industry, or other)
- Study design (masking, etc)
- Study duration
- Study location (Hospital/University or not)
- Number of participants
- Minimum age of participants
- Gender of participants

Through data exploration and a set of classification algorithms, we were able to determine that out of all of our selected features, the number of participants enrolled in the study was a driving factor in predicting whether or not a study was to make it to completion, followed by the duration of the study. We do acknowledge that there are many more trial properties that we could assess such as outcome measures and drug type categories (radiation, immunotherapy, etc). However, we believe that the framework provided by this study gives way to starting a new conversation when structuring a clinical trial that is more likely to reach completion.